

Математический подход к устранению искажений масс-спектров, зарегистрированных на склонах хроматографических пиков

***А.С. Самохин¹, А.Е. Гусев¹, Ю.А. Каламбет²**

¹Московский государственный университет имени М.В. Ломоносова,
Российская Федерация, 119991, г. Москва, Ленинские горы, 1-3

²ООО «Амперсанд», Российская Федерация, 123182, г. Москва, пл. Академика Курчатова, 2-2

Адрес для переписки: Самохин Андрей Сергеевич, E-mail: andrey.s.samokhin@gmail.com

Поступила в редакцию 17 апреля 2017 г., после исправления – 11 мая 2017 г.

Искажение масс-спектров, зарегистрированных на склонах хроматографических пиков, характерно для сканирующих масс-спектрометров (к которым относятся квадрупольные и магнитные секторные приборы). В единицу времени через сканирующий масс-анализатор могут пролетать ионы, характеризующиеся только одним единственным значением m/z . В процессе развертки масс-спектра все интересующие значения m/z регистрируются последовательно. Если за время регистрации одного масс-спектра концентрация вещества в источнике ионов существенно меняется, то полученный масс-спектр будет искажен. В работе предложен математический алгоритм, позволяющий устранять такие искажения. Алгоритм основан на восстановлении формы аналитического сигнала в пространстве между дискретными точками путем аппроксимации небольшого фрагмента масс-хроматограммы полиномом третьей степени. Число точек, используемых при аппроксимации, выбирается алгоритмом автоматически (при описании базовой линии алгоритм, как правило, использует широкие аппроксимационные окна, а при описании участков хроматограммы, содержащих пики, – узкие). Если исходные данные не могут быть адекватно аппроксимированы полиномом третьей степени (в соответствии с критерием χ^2), проводится кубическая интерполяция данных. Наряду с устранением искажений масс-спектров алгоритм сглаживает шумы в исходных данных. Эффективность алгоритма продемонстрирована на примере данных, зарегистрированных на квадрупольном масс-спектрометре.

Ключевые слова: газовая хроматография/масс-спектрометрия, ГХ/МС, сканирующий масс-анализатор, квадрупольный масс-анализатор, искажение масс-спектров, фильтрация шумов, метод Савицкого-Голея.

For citation: *Analitika i kontrol'* [Analytics and Control], 2017, vol. 21, no. 2, pp. 144-152

DOI: 10.15826/analitika.2017.21.2.007

Mathematical approach to eliminating the distortions of mass spectra recorded on the slopes of chromatographic peaks

***Andrey S. Samokhin¹, Artem E. Gusev¹, Yury A. Kalambet²**

¹Lomonosov Moscow State University, Leninskie Gory 1-3, 119991 Moscow, Russian Federation

²Ampersand Ltd., pl. Academician Kurchatov 2-2, 123182 Moscow, Russian Federation

*Corresponding author: Andrey S. Samokhin, E-mail: andrey.s.samokhin@gmail.com

Submitted 17 April 2017, received in revised form 11 May 2017

Distortions of mass spectra recorded on the slopes of chromatographic peaks (spectral skewing) are typical for the scanning instruments (quadrupoles and magnetic sectors). Only ions with the specific m/z value can pass through the scanning mass analyzer per unit time. All m/z values are scanned sequentially. If the concentration of the analyte in the ion source changes appreciably during the scan time, then the respective mass spectrum is distorted. A mathematical approach to eliminating spectral skewing is proposed

in this paper. Deskewing algorithm is based on the signal reconstruction in the space between measurement (discrete) points. To do this, a small segment of particular extracted ion chromatogram is approximated by the cubic polynomial. The size of the approximation interval is chosen automatically by the algorithm (large window sizes are typically used for the baseline and small window sizes are used for the peaks). If the data is not adequately approximated by the cubic polynomial (according to χ^2 test), then the cubic interpolation is performed. Along with the deskewing process, the algorithm provides noise smoothing of the original data. The performance of the algorithm is demonstrated on the data recorded using the quadrupole mass spectrometer.

Keywords: gas chromatography/mass spectrometry, GC/MS, scanning mass analyzer, quadrupole mass analyzer, spectral skewing, noise filtering, Savitzky-Golay filter

Введение

Проблема искажения масс-спектров, зарегистрированных на склонах хроматографических пиков, знакома любому исследователю, работающему в области газовой хроматографии/масс-спектрометрии (ГХ/МС). В англоязычной литературе это явление носит название «*spectral skewing*» [1, 2]. Искажение масс-спектров на склонах пиков характерно только для сканирующих масс-спектрометров, к которым относятся линейный квадрупольный, тройной квадрупольный и магнитный секторный приборы. Для квадрупольной ионной ловушки и времяпролетного масс-анализатора искажения масс-спектров на склонах пиков не наблюдаются.

В ряде работ авторы отмечают, что большинство ГХ/МС приборов оснащено именно квадрупольными масс-анализаторами [3, 4]. Чтобы проверить, справедливо ли это утверждение для российского парка приборов, нами был проведен анализ статей, опубликованных в журналах «Аналитика и контроль», «Журнал аналитической хи-

мии» и «Масс-спектрометрия» в период с 2012 по 2016 г. Из полученных результатов (рис. 1) можно сделать вывод, что в области ГХ/МС квадрупольные масс-анализаторы являются наиболее популярными и используются в 69 % работ. Доля сканирующих масс-спектрометров составляет 87 %.

Сканирующие масс-анализаторы представляют собой своеобразные «масс-фильтры», через которые в единицу времени могут пролетать ионы, имеющие только определенное отношение массы к заряду (m/z). Для регистрации всего масс-спектра необходимо последовательно пропустить через масс-анализатор все интересующие ионы или, другими словами, провести развертку масс-спектра. В большинстве современных квадрупольных масс-спектрометров развертку спектра проводят от минимального значения m/z к максимальному. Если в этом случае концентрация аналита в источнике ионов будет монотонно возрастать, то интенсивность масс-спектральных пиков, расположенных в области низких значений m/z спектра, будет занижена, а пиков, расположенных в области больших значений m/z , – завышена. При использовании сканирующих приборов неискаженный масс-спектр можно получить только в том случае, когда концентрация аналита в источнике ионов остается постоянной в процессе развертки масс-спектра. Именно по этой причине для аналитических целей желательно использовать масс-спектр, зарегистрированный в максимуме пика, где концентрация меняется не так быстро (или проводить усреднение масс-спектра по всему хроматографическому пику). Искажение масс-спектров на склонах хроматографических пиков, наблюдаемое при использовании сканирующих приборов, схематично проиллюстрировано на рис. 2.

Согласно литературным данным искажение масс-спектров на склонах хроматографических пиков может приводить к ряду неприятных последствий. Во-первых, существует мнение, что использование искаженных масс-спектров снижает эффективность поиска по базам данных [2]. Однако проведенные нами предварительные исследования не подтверждают это предположение. Во-вторых, ГХ/МС данные, полученные на сканирующих приборах, не могут быть корректно обработаны большинством классических хемометрических подходов, что объясняется нарушением закона аддитивности. Выполнение закона аддитивности предполагает,

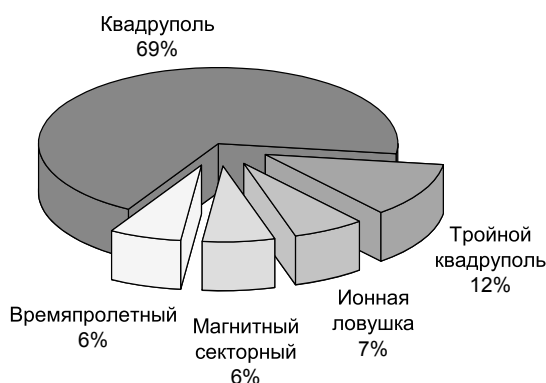


Рис. 1. Частота использования ГХ/МС приборов с масс-анализаторами различных типов. Результаты получены на основании анализа 74 работ, опубликованных в журналах «Аналитика и контроль» (13 работ), «Журнал аналитической химии» (31 работа) и «Масс-спектрометрия» (30 работ) за период с 2012 по 2016 г. Работы, опубликованные иностранными авторами, не рассматривали

Fig. 1. Popularity of GC/MS instruments equipped with various types of mass analyzers. The assessment was obtained by the analysis of 74 articles published from 2012 till 2016 in three Russian journals: "Analytical Control" (13 articles), "Journal of Analytical Chemistry" (31 articles) and "Mass Spectrometry" (30 articles). Articles of foreign authors were not considered

что в любой точке хроматографического пика чисто-го компонента отношение токов двух любых видов ионов должно оставаться постоянным. Однако, как видно из рис. 2, в случае сканирующих приборов это условие не выполняется (отношение тока иона a к току иона c увеличивается при движении сле-ва направо: $0.2 \rightarrow 1 \rightarrow 5$). Нарушение закона аддитивности объясняется тем, что всем значениям m/z , зарегистрированным в процессе развертки одного масс-спектра, приписывается одно и то же время регистрации (хотя в действительности это не так).

Искажение масс-спектров на склонах пиков, характерное для сканирующих приборов, можно устранить математически. Такой подход был опи-сан в литературе более 30 лет назад [5]. В основе подхода лежит математическое восстановление формы хроматографического сигнала в проме-жутке между соседними дискретными точками. Это позволяет рассчитать реальное значение ин-тенсивности сигнала, наблюдаемое в определен-ный момент времени на конкретной масс-хрома-

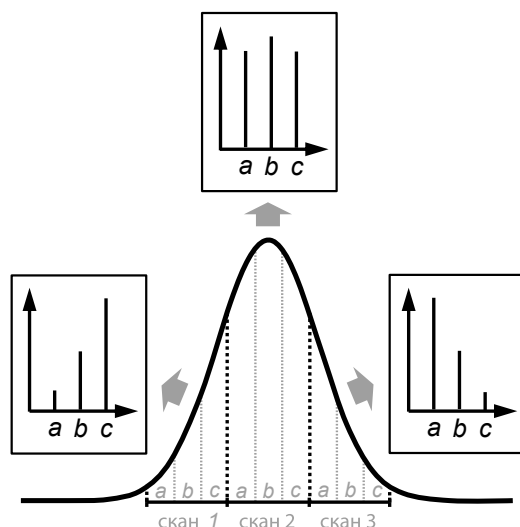


Рис. 2. Искажения масс-спектров, характерные для сканирующих приборов и наблюдаемые при бы-стром изменении концентрации аналита в источ-нике ионов. В неискаженном масс-спектре ин-тенсивности всех трех пиков (a , b , c) должны быть равны. В процессе элюирования хроматогра-фического пика зарегистрированы три масс-спе-тра: на левом склоне (скан 1), в области вершины пика (скан 2) и на правом склоне пика (скан 3). В процессе развертки масс-спектра первым реги-стрировали ион a , затем – ион b и в конце – ион c

Fig. 2. Spectral skew produced by scanning instru-ments when the concentration of the analyte in the ion source changes rapidly. All mass spectral peaks (a , b , c) should have identical abundances in the undistorted spectrum. Three mass spectra were recorded during the elution of chromatographic peak: on the left slope (scan 1), near the apex of the peak (scan 2) and on the right slope of the peak (scan 3). During the mass spectrum scanning ions were registered in the following order: ion a , then – ion b and finally ion c

тограмме, а не использовать для формирования масс-спектра ближайшее измеренное значение. Позже этот подход был использован в ряде работ [6, 7] и реализован в программном обеспечении AMDIS [8]. Следует подчеркнуть, что в работах [6–8] для расчета интенсивности сигнала в промежут-ке между дискретными точками авторы проводили интерполяцию трех последовательных точек хро-матограммы параболой. При интерполяции число степеней свободы равно нулю, поэтому интерпо-лирующая функция проходит через все исходные точки и, естественно, описывает не только полез-ный сигнал, но и шум. В связи с этим мы считаем, что работу алгоритма, устраняющего искажения масс-спектров, можно улучшить, проводя аппрок-симацию данных вместо интерполяции.

В хроматографии аппроксимация степенными полиномами традиционно применяется для сгла-живания экспериментальных данных и известна как метод Савицкого-Голея [9]. Для расчета одно-го сглаженного значения проводят аппроксимацию $(2n + 1)$ исходных точек и определяют значение по-линома в центральной точке. Число аппроксими-руемых точек или, другими словами, ширина окна аппроксимации определяет эффективность шумо-подавления. Чем шире окно, тем эффективнее по-давляются шумы, однако использование слишком широких окон приводит к искажению аналитическо-го сигнала. Ширина окна аппроксимации в методе Савицкого-Голея задается один раз и остается по-стоянной для всех точек хроматограммы. В связи с этим, если на хроматограмме присутствуют узкие пики, то невозможно добиться существенного пода-вления шума, не исказив при этом полезный сигнал.

Не так давно классический метод Савицкого-Го-лея был существенно усовершенствован [10]. Пред-ложенный алгоритм получил название адаптивный (или доверительный) фильтр. Основное отличие от метода Савицкого-Голея заключается в том, что оп-тимальная ширина окна аппроксимации (a при же-лании исследователя и оптимальная степень по-линома) подбирается алгоритмом автоматически и отдельно для каждой точки исходных данных. Опти-мальной считается та аппроксимационная модель, которая характеризуется минимальным доверитель-ным интервалом, рассчитанным для сглаженного значения интенсивности сигнала. При описании ба-зовой линии адаптивный фильтр, как правило, ис-пользует широкие окна, а при описании участков хроматограммы, содержащих пики, – узкие.

Цель настоящей работы заключалась в раз-работке математического подхода, позволяющего устранять искажения масс-спектров, зарегистри-рованных на склонах хроматографических пиков, и основанного на восстановлении формы аналитического сигнала в пространстве между дискрет-ными точками путем аппроксимации данных поли-номом третьей степени.

Экспериментальная часть

Алгоритм устранения искажений масс-спектров. Алгоритм, реализованный нами для устранения искажений масс-спектров, зарегистрированных на склонах хроматографических пиков, состоит из двух основных этапов (рис. 3): 1 – расчет истинной временной шкалы для каждой масс-хроматограммы; 2 – расчет «истинных» значений интенсивностей (путем аппроксимации данных набором полиномов

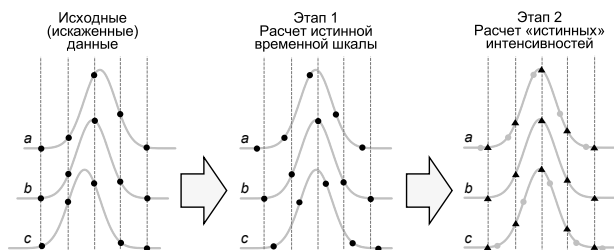


Рис. 3. Схематическое изображение ГХ/МС данных на различных этапах работы алгоритма, устраняющего искажения масс-спектров на склонах хроматографических пиков

Fig. 3. Schematic representation of GC/MS data at various stages of deskewing algorithm

третьей степени). В качестве входных параметров необходимо задать: диапазон сканирования, направление развертки масс-спектра, а также время, необходимое прибору для подготовки к регистрации следующего масс-спектра (overhead time).

На первом этапе необходимо рассчитать истинную временную шкалу отдельно для каждой масс-хроматограммы. Как видно из рис. 2, все три иона (a, b и c), зарегистрированные в пределах одного скана, отвечают различным временам удерживания. Несмотря на это, в памяти компьютера сохраняется только одно значение времени, отвечающее, как правило, середине соответствующего временного интервала (рис. 3). Рассчитать истинное время для иона, имеющего определенное значение m/z , можно по следующей формуле, которая справедлива, если развертка масс-спектра происходит линейно от минимального значения m/z к максимальному:

$$t_{\text{true}}^{(m/z=i)} = t + (t_{\text{scan}} - t_{\text{overhead}}) \cdot \frac{m/z_i - m/z_{\text{min}} + 0.5}{m/z_{\text{max}} - m/z_{\text{min}} + 1},$$

где t_{true} – истинное время регистрации иона с $m/z = i$; t – исходное (искаженное) время регистрации иона; t_{scan} – время регистрации одного масс-спектра (с учетом overhead time); t_{overhead} – overhead time; m/z_{min} и m/z_{max} – значения m/z , отвечающие началу и концу диапазона сканирования соответственно.

Расчет истинной временной шкалы приводит к сдвигу соответствующей масс-хроматограммы вдоль оси времени на небольшую величину, не превышающую $1/2$ времени регистрации одного масс-спектра (рис. 3). Чтобы устранить искажения

масс-спектров на склонах пиков, необходимо рассчитать «истинные» значения интенсивностей, достигаемые на каждой масс-хроматограмме в один и тот же момент времени (рис. 3). В настоящей работе для этой цели проводили аппроксимацию небольшого фрагмента хроматограммы полиномом третьей степени. Значение степенной функции, достигаемое в интересующий момент времени, рассматривали как «истинное» значение аналитического сигнала. Параметры аппроксимационной модели (ширина окна и смещение интересующей точки относительно центра окна) подбирали в каждом случае независимо.

В методе ГХ/МС величина дисперсии шума определяется флуктуациями ионного тока и пропорциональна числу образовавшихся ионов [11] (поскольку процесс образования каждого отдельного иона является независимым дискретным событием). Таким образом, шум в методе ГХ/МС является гетероскедастичным, и для построения аппроксимационной модели наиболее корректно использовать взвешенный метод наименьших квадратов [12]. При этом вес каждой точки будет обратно пропорционален ее интенсивности ($w_i = 1/\sigma_i^2 \sim 1/I_i$). Однако предварительные эксперименты показали, что использование взвешенного метода в конечном итоге приводит к искажению формы хроматографических пиков, обладающих небольшим отношением сигнал/шум ($S/N < 50$, здесь и далее в работе под шумом мы понимаем четырехкратное стандартное отклонение шума). По этой причине в окончательном варианте реализации алгоритма построение аппроксимационной модели проводили с помощью классического метода наименьших квадратов (что является некорректным с точки зрения статистики, однако позволяет избежать искажения аналитического сигнала).

Процедуру выбора наилучшей аппроксимационной модели проводили независимо для каждой точки определенной масс-хроматограммы. На первом этапе, используя критерий χ^2 , проводили проверку адекватности построенной модели. Еще раз отметим, что в нашем случае шум является гетероскедастичным (в отличие от оригинальной работы [10], описывающей адаптивный алгоритм фильтрации). По этой причине мы проводили расчет взвешенной (а не обычной) остаточной суммы квадратов WSS:

$$WSS = \sum w_i (Y_i - \hat{Y}_i)^2 = \sum \frac{1}{Y_i} (Y_i - \hat{Y}_i)^2,$$

где w_i – вес каждой точки, используемой при построении аппроксимационной модели ($w_i = 1/Y_i$), Y_i – исходное значение интенсивности аналитического сигнала, \hat{Y}_i – предсказанное моделью значение интенсивности.

Рассчитанное значение WSS сравнивали со значением параметра N_f (noise factor) [8] с помощью критерия χ^2 . Параметр N_f представляет собой характеристику шума ГХ/МС данных, независимую

(согласно литературным данным) от интенсивности сигнала и рассчитываемую по следующей формуле:

$$N_f^2 = \frac{s^2}{\bar{Y}} = \frac{\frac{\sum(Y_i - \bar{Y})^2}{N-1}}{\bar{Y}} = \frac{\sum \frac{1}{Y} (Y_i - \bar{Y})^2}{N-1},$$

где s^2 – дисперсия шума, рассчитанная на прямолинейном участке хроматограммы, \bar{Y} – среднее значение интенсивности сигнала на данном участке, Y_i – значение интенсивности сигнала в i -й точке, N – число точек на рассматриваемом участке хроматограммы.

В оригинальной работе, описывающей адаптивный алгоритм фильтрации [10], наилучшей считалась та аппроксимационная модель, которая удовлетворяла критерию χ^2 и обеспечивала при этом минимальный доверительный интервал. В нашем случае провести корректный расчет доверительного интервала не представляется возможным (поскольку в условиях гетероскедастичности некорректно использовать невзвешенный метод наименьших квадратов). В связи с этим, для выбора «наилучшей» модели мы использовали следующие правила (которые по своей сути близки к задаче поиска минимального доверительного интервала): 1 – вначале из всех моделей, адекватно описывающих исходные данные, отбирали только те, которые характеризовались наибольшим окном аппроксимации; 2 – затем выбирали ту модель, в которой интересующая точка имела наименьший коэффициент влияния (leverage) при невзвешенной регрессии: $h = \mathbf{x}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}$ (\mathbf{X} – матрица исходных переменных, \mathbf{x} – вектор переменных, характеризующий интересующую точку).

Следует отметить, что при описании узких (ширина у основания менее 10 точек) и интенсивных ($S/N > 200$) хроматографических пиков ни одна аппроксимационная модель не могла пройти проверку на адекватность. В этом случае для восстановления формы сигнала (в пространстве между дискретными точками) проводили интерполяцию исходных данных полиномом третьей степени.

Условия ГХ/МС анализа. Анализ толуола проводили методом капиллярной хромадистилляции в сочетании с масс-спектрометрией. В качестве детектора выступал квадрупольный масс-спектрометр «Agilent 5973» (Agilent). Температура источника ионов – 200 °С. Энергия ионизирующих электронов – 70 эВ. Диапазон сканирования от 50 до 300 а.е.м. Направление сканирования – от максимального значения m/z к минимальному.

Анализ модельного раствора стероидов проводили на квадрупольном масс-спектрометре «Trace DSQ» (Thermo Electron), соединенном с газовым хроматографом «Trace GC ultra» (Thermo Electron). Для разделения использовали капиллярную колонку Restek RTX-1 ms, 30 м × 0.32 мм × 0.25 мкм. Температура источника ионов – 250 °С. Энергия ионизирующих электронов – 70 эВ. Диапазон сканирова-

ния от 50 до 350 а.е.м. Направление сканирования – от минимального значения m/z к максимальному.

Результаты и их обсуждение

В методе ГХ/МС уровень шума зависит от интенсивности сигнала. Согласно литературным данным [8], отношение стандартного отклонения шума к квадратному корню из интенсивности сигнала (параметр N_f) является характеристикой конкретного прибора и не меняется с ростом интенсивности или при переходе от одной масс-хроматограммы к другой. Проверка адекватности аппроксимационной модели, реализованная в нашем алгоритме, базируется на данном утверждении. Чтобы удостовериться в постоянстве параметра N_f нами были проанализированы экспериментальные данные, полученные методом капиллярной хромадистилляции в сочетании с масс-спектрометрией [13, 14]. В методе хромадистилляции чистое органическое вещество элюируется из системы в виде ступени, высота которой в ряде случаев остается постоянной на всем ее протяжении. На рис. 4 представлено

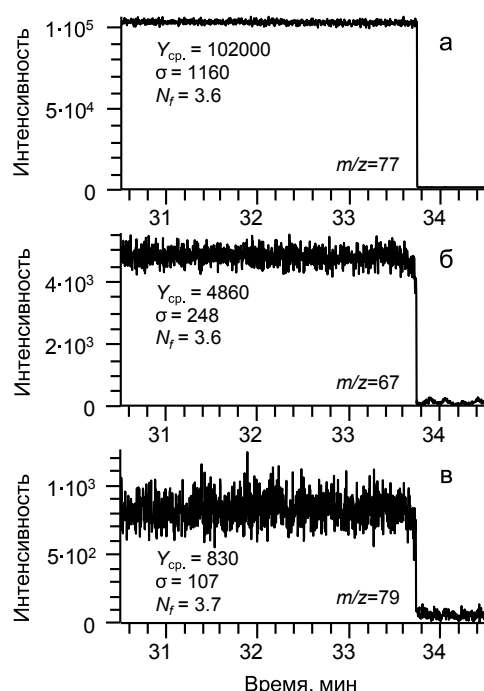


Рис. 4. Фрагмент ступени толуола, зарегистрированной методом капиллярной хромадистилляции/масс-спектрометрии для трех значений m/z : 77 (а), 67 (б) и 79 (в). Для верхнего плато ступени (31.0–33.0 мин) рассчитано среднее значение интенсивности сигнала ($Y_{cp.}$), стандартное отклонение (σ) и параметр N_f

Fig. 4. Segment of elution profile of toluene (having the form of a step) obtained using the capillary chromadistillation coupled with mass spectrometry for three m/z values: 77 (a), 67 (b) and 79 (v). Average value of the signal ($Y_{cp.}$), standard deviation (σ) and noise factor (N_f) were calculated for the upper plateau (31.0–33.0 min)

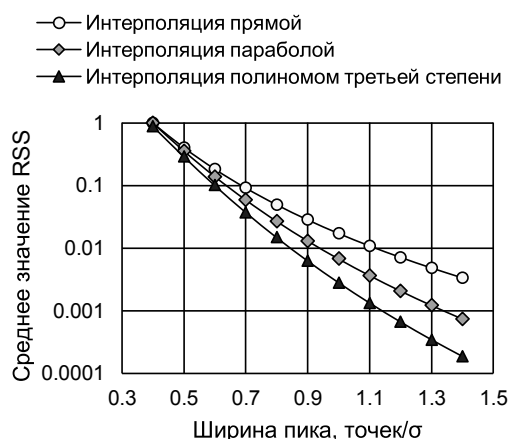


Рис. 5. Среднее (по 10 различным дискретным пикам) значение величины остаточной суммы квадратов (RSS), рассчитанное между истинным профилем кривой Гаусса ($\pm 4\sigma$) и профилем пика, восстановленным при помощи интерполяции. Значение RSS нормировано на максимальное значение

Fig. 5. Average (for 10 various discrete peaks) residual sum of squares (RSS) calculated between true Gauss curve ($\pm 4\sigma$) and its profile reconstructed by the interpolation. RSS was normalized to the maximum value

несколько хроматодистграмм, содержащих ступень толуола и отвечающих различным значениям m/z . Для прямолинейного участка ступени был проведен расчет средней интенсивности сигнала, стандартного отклонения шума и значения параметра N_f . В общей сложности было рассмотрено 26 значений m/z , присутствующих в масс-спектре электронной ионизации толуола. Следует отметить, что наиболее интенсивные значения m/z (91, 92) не принимали во внимание, поскольку по ним наблюда-

лось насыщение детектора. Значение параметра N_f (рассчитанное для различных значений m/z и интенсивностей сигнала) находилось в диапазоне от 3.3 до 4.1. Полученные результаты подтверждают хорошую воспроизводимость параметра N_f и позволяют использовать данный параметр при проверке адекватности модели по критерию χ^2 .

Как было отмечено в экспериментальной части, если ни одна аппроксимационная модель не описывала исходные данные адекватно (согласно критерию χ^2), то для расчета «истинного» значения интенсивности проводили интерполяцию данных. В отличие от [6–8] в качестве интерполирующей функции использовали полином третьей (а не второй) степени. Это объясняется тем, что согласно нашим исследованиям, полином третьей степени описывает форму кривой Гаусса (в области $\pm 4\sigma$) с меньшей ошибкой, что проиллюстрировано на рис. 5.

Следует подчеркнуть, что поскольку для восстановления формы сигнала в пространстве между точками мы (по возможности) проводили аппроксимацию данных, то уровень шума в «исправленных» ГХ/МС данных существенно снижался. В связи с этим, разработанный алгоритм можно использовать не только для устранения искажений масс-спектров, зарегистрированных на склонах пиков, но и просто для фильтрации шумов «неискаженных» ГХ/МС данных (зарегистрированных, например, при использовании ионной ловушки или времяпролетного масс-спектрометра).

На рис. 6 представлены два фрагмента масс-хроматограммы ($m/z = 129$), зарегистрированные при проведении анализа модельного раствора стероидов.

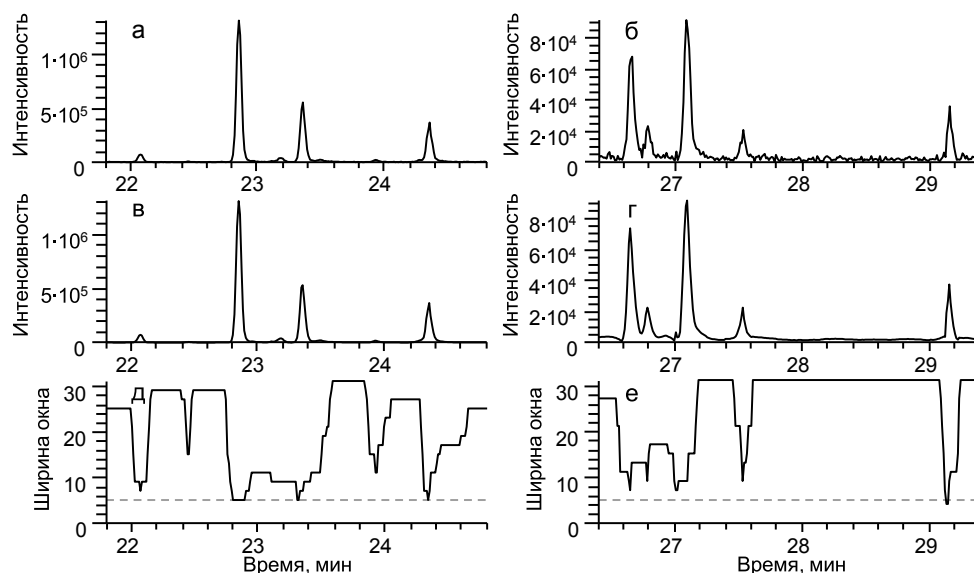


Рис. 6. Два исходных (а, б) и два сглаженных (в, г) фрагмента масс-хроматограммы ($m/z = 129$), зарегистрированные при проведении анализа модельного раствора стероидов, а также ширина окна аппроксимации, выбранная алгоритмом для каждой точки исходных данных (д, е)

Fig. 6. Two segments of extracted ion chromatogram ($m/z = 129$) of a model mixture of steroids corresponding to the initial (a, б) and smoothed data (в, г), as well as the size of approximation interval (д, е) automatically chosen by the algorithm for each point of the initial data

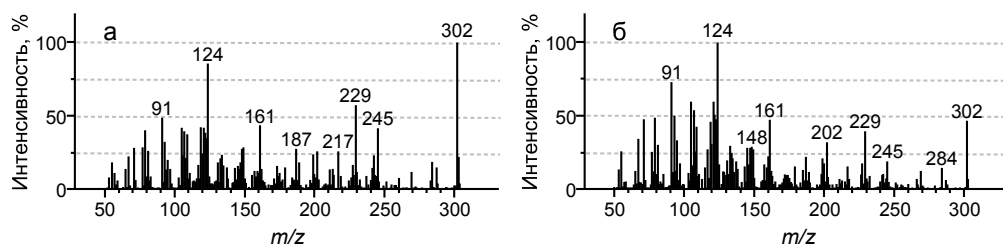


Рис. 7. Масс-спектры метилтестостерона, зарегистрированные на левом (а) и правом (б) склонах хроматографического пика при обработке исходных (искаженных) ГХ/МС данных

Fig. 7. Mass spectra of methyltestosterone recorded on the left (a) and right (b) slopes of the respective chromatographic peak. Initial (distorted) GC/MS data were considered

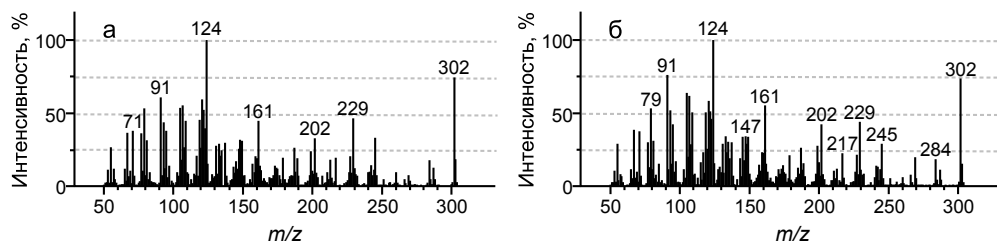


Рис. 8. Масс-спектры метилтестостерона, зарегистрированные на левом (а) и правом (б) склонах хроматографического пика при обработке исправленных ГХ/МС данных

Fig. 8. Mass spectra of methyltestosterone recorded on the left (a) and right (b) slopes of the respective chromatographic peak. Deskewed GC/MS data were considered

Один фрагмент содержит интенсивные ($S/N = 300$ для пика с $t_R = 22.84$ мин), а другой, наоборот, малоинтенсивные пики ($S/N < 25$). Оба фрагмента были «сглажены» разработанным нами алгоритмом (расчет истинной временной шкалы в данном случае не проводили). Из рисунка видно, что в результате работы алгоритма величина шума сократилась в несколько раз, а форма пиков при этом не искажалась. Из представленных на рис. 6 графиков видно, что при описании базовой линии алгоритм использует широкие аппроксимационные окна, вплоть до максимального значения, равного в этой работе 31. При описании пиков алгоритм, наоборот, строит аппроксимационную модель с использованием небольшого числа точек (от 5 до 9). Следует отметить, что в некоторых случаях вершина пика не может быть адекватно описана ни одной аппроксимационной моделью (например, это наблюдается в случае пика с $t_R = 29.16$ мин).

На рис. 7 представлены масс-спектры метилтестостерона, зарегистрированные на склонах соответствующего хроматографического пика (ширина пика на 10 % высоты составляет ~13 точек). Искажения масс-спектров, обусловленные сканирующей природой масс-спектрометра, видны невооруженным глазом: пик молекулярного иона ($m/z = 302$) является основным пиком в масс-спектре, зарегистрированном на левом склоне, однако его интенсивность в масс-спектре, зарегистрированном на правом склоне, составляет всего 45 %. Обработка исходных ГХ/МС данных разработанным нами алгоритмом позволила устранить искажения масс-спектров, что проиллюстрировано на рис. 8. На рис. 9а показано, как меняется отноше-

ние интенсивностей двух масс-спектральных пиков ($m/z = 124$ и $m/z = 302$) при переходе от одной точки хроматографического пика к другой. Из представленного графика видно, что в случае исходных (искаженных) ГХ/МС данных отношение $I_{m/z=124} / I_{m/z=302}$ возрастает при движении по профилю пика слева направо. Для исправленных ГХ/МС данных отношение $I_{m/z=124} / I_{m/z=302}$ практически не меняется.

Статистический анализ большого массива экспериментальных данных, проведенный нами в рамках настоящей работы, показывает, что отношение интенсивностей двух масс-спектральных пиков (наблюдаемое в определенной точке хроматограммы при отсутствии систематических искажений масс-спектров) с вероятностью 95 % находится в диапазоне от $\bar{r} - 1.87 \cdot s$ до $\bar{r} + 1.99 \cdot s$, где \bar{r} и s – среднее значение и стандартное отклонение отношения интенсивностей рассматриваемой пары масс-спектральных пиков. Пунктирные горизонтальные линии на рис. 9 представляют собой границы 95 % доверительного интервала для отношения $I_{m/z=124} / I_{m/z=302}$. Как видно из рис. 9а, в случае исправленных ГХ/МС данных все точки располагаются внутри 95 % доверительного интервала, хотя для исходных ГХ/МС данных таких точек было всего 3.

В рассматриваемом примере ширина хроматографического пика метилтестостерона составляет ~19 точек у основания (или около 13 точек на 10 % высоты). Увеличение скорости сканирования позволяет уменьшить искажения масс-спектров, зарегистрированных на склонах хроматографических пиков, без использования математических алгоритмов. При достижении определенного числа точек на пик систематической погрешностью величины

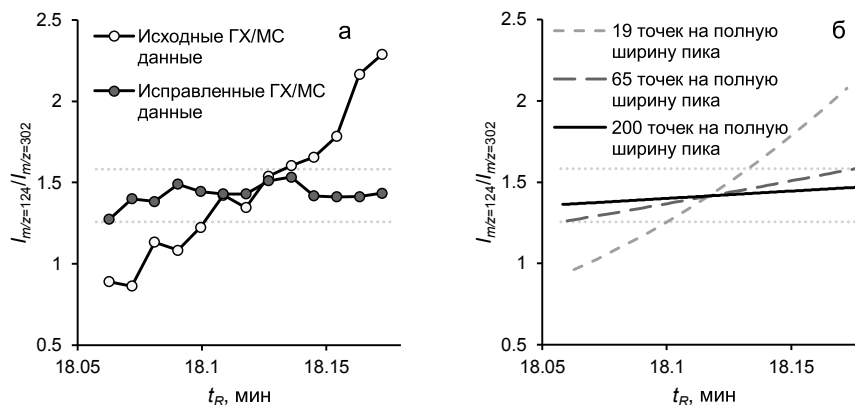


Рис. 9. Отношение интенсивностей двух масс-спектральных пиков ($I_{m/z=124}/I_{m/z=302}$), присутствующих в масс-спектре метилтестостерона, наблюдаемое в различных точках соответствующего хроматографического пика. Результаты реального эксперимента представлены слева (а), результаты модельного компьютерного эксперимента представлены справа (б). Пунктирные горизонтальные линии – границы 95 % доверительного интервала. Отношение $I_{m/z=124}/I_{m/z=302}$ рассчитано только для тех точек хроматографического пика, интенсивность которых превышает 10 %

Fig. 9. Ratio of two mass spectral peaks (presented in spectrum of methyltestosterone) calculated for the different points of respective chromatographic peak. Real experimental data are shown in the left plot (a), data obtained by the computer experiment are shown in the right plot (b). Dashed horizontal lines indicate 95 % confidence interval. Ratio $I_{m/z=124}/I_{m/z=302}$ is calculated only for those points of the chromatographic peak which intensity exceeds 10 %

$I_{m/z=124}/I_{m/z=302}$ можно будет пренебречь, поскольку она не будет превышать 1/3 доверительного интервала. Проведенный нами компьютерный модельный эксперимент показывает, что это наблюдается, когда число точек у основания достигает 200 (рис. 9б). Следует подчеркнуть, что при расчете отношения $I_{m/z=124}/I_{m/z=302}$ мы рассматриваем только те точки хроматографического пика, интенсивность которых превышает 10 %, поскольку искажения масс-спектров проявляются тем сильнее, чем ближе расположен масс-спектр к началу или концу хроматографического пика. Современные квадрупольные масс-спектрометры позволяют регистрировать данные со скоростью 40–50 спектров/с (если диапазон сканирования составляет 50–350 единиц шкалы m/z). В рассматриваемом примере ширина хроматографического пика у основания должна составлять как минимум 4 с, чтобы достичь указанного числа точек на пик. Таким образом, математическое устранение искажений масс-спектров остается актуальным для узких хроматографических пиков. Разработанный алгоритм представляет особый интерес для исследователей, использующих метод двумерной газовой хроматографии, в котором ширина хроматографического пика у основания не превышает 0.5 с.

В заключении следует отметить, что при использовании разработанного алгоритма следует уделять особое внимание процессу калибровки масс-спектрометра. Например, некорректная установка «нуля» детектора может привести к тому, что шум базовой линии будет частично «обрезан», что не позволит рассчитать параметр N_f и использовать

критерий χ^2 для проверки адекватности построенных аппроксимационных моделей. Другая возможная проблема заключается в том, что в процессе загрязнения источника ионов на масс-хроматограммах могут появляться единичные всплески или, наоборот, провалы сигнала, интенсивность которых превышает в несколько раз стандартное отклонение шума. Такие резкие скачки интенсивности не могут быть сглажены предлагаемым алгоритмом, и для борьбы с ними следует использовать специальные математические подходы.

Заключение

В работе предложен подход к устранению искажений масс-спектров, зарегистрированных на склонах хроматографических пиков. Подход основан на восстановлении формы аналитического сигнала в пространстве между дискретными точками путем аппроксимации данных полиномом третьей степени. Эффективность алгоритма продемонстрирована на примере реальных ГХ/МС данных, зарегистрированных на квадрупольном масс-спектрометре.

Благодарности

Авторы работы выражают благодарность Чепелянскому Д.А. (химический факультет МГУ) за предоставление экспериментальных данных, полученных методом капиллярной хроматографии в сочетании с масс-спектрометрией.

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 16-33-60169 мол_а_дк.

Acknowledgements

We are grateful to D.A. Chepelyansky (Chemistry Department of Lomonosov Moscow State University) for providing experimental data recorded using the capillary chromatodistillation coupled with mass spectrometry.

The reported study was funded by The Russian Foundation for Basic Research, according to the research project No. 16-33-60169 mol_a_dk

Литература

1. Smith R. Understanding mass spectra: a basic approach. New York: Wiley, 2004. 372 p.
2. Watson T., Sparkman D. Introduction to mass spectrometry: instrumentation, applications, and strategies for data interpretation. Chichester: Wiley, 2008. 862 p.
3. Cazes J. Encyclopedia of chromatography, Volume 1. Boca Raton: CRC Press, 2005. 800 p.
4. Bouchonnet S. Introduction to GC-MS coupling. Boca Raton: CRC Press, 2013. 228 p.
5. King M.D., King G.S. A numerical method for extracting mass spectra from gas chromatography mass spectrometry data arrays // *Anal. Chem.* 1985. V. 57, № 6. P. 1049-1056.
6. Pool W.G., de Leeuw J.W., van de Graaf B. A rapid routine to correct for skewing in gas chromatography/mass spectrometry // *J. Mass Spectrom.* 1996. V. 31, № 2. P. 213-215.
7. Fraga C.G. Chemometric approach for the resolution and quantification of unresolved peaks in gas chromatography-selected-ion mass spectrometry data // *J. Chromatogr. A.* 2003. V. 1019, № 1-2. P. 31-42.
8. Stein S.E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data // *J. Am. Soc. Mass Spectrom.* 1999. V. 10, № 8. P. 770-781.
9. Savitzky A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures // *Anal. Chem.* 1964. V. 36, № 8. P. 1627-1639.
10. Каламбет Ю., Мальцев С., Козьмин Ю. Фильтрация шумов: окончательное решение проблемы // *Аналитика*. 2011. Т. 1. С. 50-55.
11. Signal-to-noise ratios in the measurement of low ion currents using electron multipliers / F.M. Harris [et al.] // *Mass Spectrom. Rev.* 1984. V. 3, № 2. P. 209-229.
12. Каламбет Ю.А., Мальцев С.А., Козьмин Ю.П. Доверительные интервалы метода взвешенных наименьших квадратов и стратегия градуировки // *Заводская лаборатория. Диагностика материалов*. 2015. Т. 81, № 1. С. 69-76.
13. Zhukhovitskii A.A., Yanovskii S.M., Shvarzman V.P. Chromatodistillation // *J. Chromatogr.* 1976. V. 119, P. 591-598.
14. Капиллярная хромадистилляция в сочетании с масс-спектрометрией и определение примесей в органических растворах / И.В. Гуляев [и др.] // *Масс-спектрометрия*. 2012. Т. 9, № 2. С. 117-120.

References

1. Smith R. *Understanding mass spectra: a basic approach*. New York, Wiley, 2004. 372 p. DOI: 10.1002/0471479357
2. Watson T., Sparkman D. *Introduction to mass spectrometry: instrumentation, applications, and strategies for data interpretation*. Chichester, Wiley, 2008. 862 p. DOI: 10.1002/9780470516898
3. Cazes J. *Encyclopedia of chromatography, Volume 1*. Boca Raton, CRC Press, 2005. 800 p.
4. Bouchonnet S. *Introduction to GC-MS coupling*. Boca Raton, CRC Press, 2013. 228 p.
5. King M.D., King G.S. A numerical method for extracting mass spectra from gas chromatography mass spectrometry data arrays. *Anal. Chem.*, 1985, vol. 57, no. 6, pp. 1049-1056. DOI: 10.1021/ac00283a020
6. Pool W.G., de Leeuw J.W., van de Graaf B. A rapid routine to correct for skewing in gas chromatography/mass spectrometry. *J. Mass Spectrom.*, 1996, vol. 31, no. 2, pp. 213-215. DOI: 10.1002/(SICI)1096-9888(199602)31:2<213::AID-JMS284>3.0.CO;2-6
7. Fraga C.G. Chemometric approach for the resolution and quantification of unresolved peaks in gas chromatography-selected-ion mass spectrometry data. *J. Chromatogr. A*, 2003, vol. 1019, no. 1-2, pp. 31-42. DOI: 10.1016/S0021-9673(03)01329-3
8. Stein S.E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.*, 1999, vol. 10, no. 8, pp. 770-781. DOI: 10.1016/S1044-0305(99)00047-1
9. Savitzky A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 1964, vol. 36, no. 8, pp. 1627-1639. DOI: 10.1021/ac60214a047
10. Kalambet, Y., Maltsev, S., Kozmin, Y. [Noise filtering: the final solution]. *Analitika [Analytics]*, 2011, vol. 1, pp. 50-55 (in Russian).
11. Harris F.M., Trott G.W., Morgan T.G., Brenton A.G., Kingston E.E., Beynon J.H. Signal-to-noise ratios in the measurement of low ion currents using electron multipliers. *Mass Spectrom. Rev.*, 1984, vol. 3, no. 2, pp. 209-229. DOI: 10.1002/mas.1280030203
12. Kalambet, Y.A., Maltsev, S.A., Kozmin, Y.P. [Confidence intervals for weighted least squares technique and calibration strategy]. *Zavodskaya laboratoriya. Diagnostika materialov [Industrial laboratory. Diagnostics of Materials]*, vol. 81, no. 1, pp. 69-76 (in Russian).
13. Zhukhovitskii A.A., Yanovskii S.M., Shvarzman V.P. Chromatodistillation. *J. Chromatogr.*, 1976, vol. 119, pp. 591-598. DOI: 10.1016/S0021-9673(00)86820-X
14. Gulyaev I.V., Chepelyansky D.A., Revelsky I.A., Revelsky A.I. [Capillary chromatodistillation coupled with mass spectrometry and determination of impurities in organic solutions]. *Mass-spektrometriya [Mass spectrometry]*, 2012, vol. 9, no. 2, pp. 117-120 (in Russian).